

THE EFFECT OF POSITIVE AND NEGATIVE
ADMINISTRATIONS ON INTELLIGENCE
TEST PERFORMANCE

By
WILLIAM GEORGE MURDY, JR.

A DISSERTATION PRESENTED TO THE GRADUATE COUNCIL OF
THE UNIVERSITY OF FLORIDA
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE
DEGREE OF DOCTOR OF PHILOSOPHY

UNIVERSITY OF FLORIDA

February, 1962

UNIVERSITY OF FLORIDA



3 1262 08552 2257

ACKNOWLEDGMENTS

The author wishes to express his appreciation of the guidance and stimulation received from his committee chairman, Dr. Richard J. Anderson, in the development of this research. A special note of thanks is also extended to the members of his original supervisory committee-- Dr. Dorothy A. Rethlingshafer, Dr. James C. Dixon, Dr. Malcolm H. Robertson, and Dr. Albert M. Barrett--for their interest and efforts concerning this dissertation.

Without the help of the following graduate students who served as examiners the dissertation would not have been completed: Andrew Farinacci, Donald Hartsough, Irwin Farbman, Alan Rusnak, William Coulter, Hal Rosen, James Alsobrook, and Ronald Baxter.

And, of course, the author is especially grateful for the help and reinforcement from his wife, Louise.

TABLE OF CONTENTS

	Page
ACKNOWLEDGMENTS	ii
LIST OF TABLES	iv
Chapter	
I. INTRODUCTION	1
Problem	1
Development of the Problem	2
Research with Projective Tests	4
Research with Non-Projective Tests	8
Hypotheses	11
II. EXPERIMENTAL PROCEDURE	12
Subjects	12
Examiners	13
Materials	13
Negative and Positive Treatment Conditions	17
Procedure	19
III. RESULTS AND DISCUSSION	22
Results	22
Discussion	29
IV. SUMMARY	33
BIBLIOGRAPHY	35
BIOGRAPHICAL SKETCH	38

LIST OF TABLES

Table	Page
1. Rating Scale	15
2. Analysis of Variance	26
3. Estimated Average Full Scale IQ's for Positive and Negative Treatments	26
4. Individual Subtest Scale Scores Under Each Treatment Condition	27
5. Comparison of Subjects' Ratings of Examiners Under Positive and Negative Treatments	28

CHAPTER I

INTRODUCTION

Problem

The purpose of this study was to determine whether or not the experimentally established attitudes of the examiner during the administration of an intelligence test could influence the subject's test performance. More specifically, the experimenter wished to evaluate the effects of a Positive Administration (i.e., an administration characterized by a warm and highly approving, interested manner of test administration) and of a Negative Administration (i.e., an administration characterized by a persistently rejecting and disinterested manner of test administration).

The study was concerned mainly with the examiner variable in the psychologist-subject relationship in intelligence testing. The experimenter hoped to amplify the scant amount of research in this area by demonstrating a significant difference in the subject's intelligence test performance as a result of the administration experience of two artificial, but possible, conditions of the examiner attitude.

Moreover, in order to understand more fully the interaction between the examiner and the subject, the experimenter was interested in identifying, through the use of a rating scale, aspects of examiner performance which are present in one form of administration and not present in the other.

Development of the Problem

Consideration of interpersonal and situational variables in psychological testing has been a matter of major concern for about a dozen years. Joel (1949) cogently states:

It has been argued that the effect of the examiner should be reduced to a minimum by assuming a constant warm attitude which does not change under the impact of the subject's personality. There can be no argument about the desirability of a friendly attitude on the part of the examiner or about its beneficial effect on rapport. But we make a fundamental mistake if we believe the assumed attitude of the examiner can really nullify the dynamics of the testing situation. Clinical psychologists are human, and the assumed attitude of warmth notwithstanding, they react to different subjects in different ways, partly because of irrational attitudes. So we must reckon with the effect of the examiner's actual continuously changing feelings underneath the assumed attitudes.

More recently, this matter is again touched upon by Cronbach (1960, p. 60), who writes:

The tester has been accustomed to think of himself as an unemotional, impartial task-setter. His traditions encourage the idea that he, like the physical scientist or engineer, is "measuring an object" with a technical tool. But the "object" before him is a person, and testing involves a complex psychological relationship. The traditional concern with motivation and rapport recognizes this fact but . . . leads to little more than a recommendation that the tester be pleasant and encouraging, and help the subject understand the value of the test. This, we are beginning to suspect, barely touches the real social-psychological complexities of testing.

A recent review of the literature dealing with the influence of situational and interpersonal variables in projective testing (Masling, 1960) cites over seventy references dealing with the method of administration, the testing situation, examiner influence, and subject influence. Generally speaking, the crucial element in modifying the subject's projective response was the extent to which the subject's attitude toward the total testing situation was influenced by the

experimental conditions. When the experimental variable was peripheral to the examination (i.e., when it occurred before the actual test administration), no appreciable effect was introduced in the protocol.

It is necessary, before reviewing recent pertinent literature in this area, to discuss the nature of the task as a significant variable in demonstrating the effect of the testing conditions on the subject's performance. Projective tests are commonly regarded as unstructured and ambiguous stimuli. Those tests which are characterized by specific instructions, clearly defined stimuli, and right and wrong answers are regarded as structured or non-projective tests. The ambiguity of the projective testing situation enhances the probability that the examiner and the subject will be influenced by each other in completing the testing task.

Masling (1960) summarizes the idea that the subject or the examiner, in an unstructured situation, will utilize all the possible cues by stating:

The S in the projective test setting will not only use those cues furnished by the ink blot or picture, but also those supplied by his feelings about the examiner, those furnished by his needs, attitudes and fears, those implied in the instructions, the room, and previous knowledge of the test, and those cues supplied consciously or unconsciously by E. When E faces the ambiguous situation of supplying meaning to a series of isolated, discrete responses, he will not only rely on S's responses, but also on those cues furnished by his training and theoretical orientation, his own needs and expectations, his feelings about S and the constructions he places on S's test behavior and attitudes.

The implication here is that, with well structured so-called non-projective tests, the probability that situational and interpersonal variables will influence test performance and results is greatly diminished. In other words, Masling implies that because the testing

situation is well structured, the ability of the examiner to be objective and not involved personally with the subject is greatly enhanced. The following review of the literature suggests that this implication is questionable.

The amount of research on the examiner-subject relationship in intelligence testing is small. The few articles reviewed suggest, however, that this relationship is important to the subject's score and the evaluation of results. The research also suggests that the effect of this relationship varies in intensity with different groups. Poorly-adjusted people are more affected than well-adjusted people. Further, there is the indication that the socio-economic level of the subject, his cultural background, and his present life situation are all variables to be considered in evaluating the results of intelligence tests.

In view of the above development, the author felt that this experiment would be a significant contribution to the study of the influence of situational and interpersonal variables in intelligence testing.

Research with Projective Tests

The following review of representative studies of the influence of psychologist-subject relationship in projective testing will demonstrate that this relationship can modify test results.

Lord (1950) used three different atmospheres of Rorschach administration: (1) a usual standardized administration (2) a standardized administration with negative emotional loading in which the examiner assumed the role of a cold and harsh figure demonstrating no

concern for the subject, and (3) a standardized administration with positive emotional loading in which the examiner was warm and charming. Lord's method was to employ three female examiners and have each subject take the Rorschach three times (once with each examiner under a different atmosphere). She not only found that the different methods of interaction varied the protocols elicited from the same subjects, but also that a greater number of differences occurred when the examiners gave the test in their normal manner than when they gave the test under conditions of assumed rapport. Lord concluded that the underlying personality of the examiner influences the subject's Rorschach test performance to a greater extent than any assumed rapport.

Working with a similar examiner variable of negative and positive rapport and with the Rorschach test, Luft (1953) found that the subjects treated in a warm fashion indicated that they liked a significantly greater number of inkblots than those subjects treated in a cold manner.

Sanders and Cleveland (1953) approached the problem of examiner influence on Rorschach scores not by varying the mode of administration, as the previous two studies did, but by training nine male graduate students, unsophisticated as to projective techniques, to administer the Rorschach test after the experimenters obtained a personal Rorschach from each. The examiners then administered twenty Rorschachs each to undergraduate subjects who, in turn, rated the examiners on measures of overt anxiety and hostility; the examiner's covert anxiety and hostility were obtained from their Rorschach protocols. On the basis of these measures, Sanders and Cleveland found that the examiners rated in a particular way (with regard to hostility or anxiety) tended to elicit

Rorschachs which differed significantly from those Rorschachs elicited by examiners with different ratings. In short, different examiners elicited different Rorschach scores from their subjects.

Exploring the possibility of finding significant over-all differences in Rorschach protocols obtained by various examiners with similar backgrounds from a homogeneous group of patients (white, male veterans, 25-32 years old with functional rather than organic ailments), Gibbey, Miller and Walker (1953) analyzed the obtained Rorschach protocols for examiner influence and found that the examiners differed from each other significantly in the determinants they elicited from comparable subjects.

That differential instructions can influence Rorschach scores is clearly demonstrated by Henry and Rotter (1956), who gave the control group a standard administration, but informed the experimental group that the test is used to discover serious emotional disturbances. The experimental group gave significantly more cautious and conforming responses than the control group.

Whether or not the examiner is present during the testing appears to be a significant variable in influencing test performance. Bernstein (1956), using the TAT under conditions of examiner present and then absent for both written and oral TAT productions, found that the only significant difference in the stories was a function of the examiner being present or absent. The results indicated that the examiner's presence acts as an inhibiting factor for strongly emotional material.

Not only does the examiner appear to have a significant influence on projective test performance, but also the physical surroundings seem

to have an influence. Rabin, Nelson, and Clark (1954), in order to study the effects of the physical surroundings on Rorschach performance, used two experimental groups of males in differently decorated waiting rooms and a control group in an undecorated waiting room. One experimental group waited in a room decorated with anatomical charts and surgical pictures; the other group waited in a room decorated with photographs of nude and seminude females. No significant difference between groups in the number of anatomical responses was found, but there was a significant difference in the number of sexual responses. A further interesting finding of this study was that those subjects who waited in the room decorated with pictures of nude women gave significantly more sexual responses to the male examiner than to the female examiner.

Two studies using operant conditioning of the subject's verbal behavior further demonstrate that cues given by the examiner can affect test results. Wickes (1956) used a control group and two experimental groups for a series of homemade inkblots similar to the Rorschach test. Of the two experimental groups, one received verbal reinforcement (e.g., "fine" and "good") to movement responses, the other received postural reinforcement (e.g., nodding and smiling) to movement responses. By introducing reinforcement during the second half of the inkblot series, Wickes found a significant increase in movement responses for both reinforcement groups, but found no such increase in the control group. Gross (1959) in a similar study reinforced human content responses on the Rorschach, verbally to one experimental group, and posturally or non-verbally to another. It was found that both the verbally reinforced

and non-verbally reinforced groups gave significantly more human responses than the control group and that there was no significant difference between the two types of reinforcement.

The last study to be reviewed in the area of the influence of interpersonal and situational variables on projective test performance deals with the influence of the subject upon the examiner. Masling (1957) used two female subjects who acted as confederates by posing as subjects and acting warm or cold to a group of naïve examiners who had the task of interpreting sentence completion protocols yielded by the subjects. Masling found that when the subject acted warm to the examiner her protocol was interpreted more favorably (i.e., she appeared in better mental health) than when she acted coldly.

Research with Non-Projective Tests

All the pertinent research in the area of measuring the examiner influence on "structured non-projective" testing has been done with individually administered intelligence tests. The amount of research in this area is scant when compared with similar investigations in projective testing.

Intelligence tests are regarded as objective and impersonal instruments for the study of behavior. The stimuli are clearly defined; the answers are right or wrong; the instructions are specific, and the examiner reads the questions as they appear in a manual; the answers are scored according to the manual. The following review will, however, demonstrate that there is reason to suspect the objectivity of intelligence tests.

Lantz (1945) studied the effects of situational variables (i.e., success or failure) upon intelligence test performance. Using nine selected tasks from Form L of the Revised Stanford-Binet and nine comparable tasks from Form M as measures of intelligence, Lantz found that his "failure group" of subjects--who experienced failure in a game in which the task was to secure, in three different ways, a ball from a box--experienced a lack of the expected average test-retest increase in score and a decrease in the number of correct responses on those test items which involved the use of thought processes. The "failure-group" also demonstrated a decrease on the ratings of willingness, self-confidence and attention. The "success-group," however, demonstrated an increase of average scores on the mental tasks with a decrease in score variability.

Getting closer to the examiner-subject relationship, Staudt (1948) gave three groups of college students tests of verbal analogies, arithmetic, and cancellation. The first group was tested under normal conditions. The second group was instructed to work accurately. The third group was instructed to work very rapidly and was subjected to tension-evoking conditions (e.g., a buzzer sounded every thirty seconds, at which times the examiner stated how many items should have been completed). These conditions produced feelings of inferiority in the subjects, because the level of attainment was set beyond their capacities. Staudt's results demonstrated significantly more errors for the tension group and no difference between the other two groups.

Hutt (1947) compared IQ ratings obtained by "consecutive" testing (i.e., normal procedure) and "adaptive" testing (i.e., procedure of

following every failure by an easier item) on Form L of the Revised Stanford-Binet. The results showed that well-adjusted children make similar scores with the two methods of administration, whereas poorly-adjusted children make higher IQ ratings with the "adaptive" testing. The suggestion here is that, because some children experience an increasing sense of failure with the "consecutive" testing, they may therefore be unable to succeed on later test items. Hutt felt that the "adaptive" testing yielded more valid IQs with the poorly-adjusted children.

Directly manipulating the nature of the social relationship between the examiner and the subject and determining the effect of such upon intelligence test performance by establishing (1) a good relationship, (2) a poor relationship, and (3) a control group (i.e., no relationship), respectively, with three groups of nursery school children, Sacks (1952) found in all three groups a mean increase in IQ from Form L of the Revised Stanford-Binet to Form M. The children in the "relationship" groups demonstrated a significantly greater change than did those in the control group.

Masling (1959), investigating the effects of the subject upon the examiner's administration and scoring of selected subtests from the Wechsler-Bellevue II, had two subjects act either warm or cold to naïve examiners who were sophisticated in intelligence testing. The subjects also gave memorized responses, most of which were devised to be difficult to score. The study thus compared the examiner's treatment of warm and cold subjects, and the results showed conclusively that to the warm subjects, the examiner was more lenient, used more reinforcing statements, and gave more opportunity to clarify or correct responses.

Hypotheses

In light of the above research, the specific hypotheses which were tested were as follows:

- I. There will be a consistent shift in intelligence test performance as a result of the different experimentally established negative and positive interactions between the examiner and the subject.
- II. This shift will most likely reflect a decrement in score for the majority of the subjects under the negative treatment condition.
- III. The subject's ratings of the examiners will be clearly different as a result of the different treatment conditions.
- IV. The subjects will perceive the negative treatment condition examiners in an unfavorable light, whereas they will perceive the positive treatment condition examiners in a favorable light.

CHAPTER II

EXPERIMENTAL PROCEDURE

Subjects

The subjects used in this experiment were 48 male undergraduate students at the University of Florida. They acted as their own controls and were tested in both treatments yielding a total of 96 scores for comparison. Most of them were sophomores and juniors and were selected from a course in general psychology.

For purposes of sampling a wide range of intellectual ability, the A.C.E. was used as a means of selecting subjects. The wide range was developed by selecting the subjects on the basis of divisions of ranges of percentile ranks: one-sixth of the subject population were drawn from range 24 and below; one-sixth from 25 to 39; one-sixth from 40 to 54; one-sixth from 55 to 69; one-sixth from 70 to 84; one-sixth from 85 and above. This division meant that there should have been 8 subjects within each range.

Because 5 subjects did not appear and because the source of alternate subjects was limited, the actual number of subjects per division worked out as follows: 7 subjects from range 24 and below; 6 from 25 to 39; 7 from 40 to 54; 10 from 55 to 69; 9 from 70 to 84; and 9 from 85 and above.

As will be seen in the Procedure section of this chapter, each subject followed a unique pattern of testing. It was essential that

the subject follow his pattern to completion for meaningful data to occur. All of the subjects cooperated except one who dropped out before he completed his pattern. Minor modifications in the statistical treatment of the data were made and will be discussed as they are encountered.

Examiners

Hammond (1954) criticizes studies on the examiner effect in psychological testing on the basis that they fail to take an adequate sample of the examiner population and thus limit the degree to which results can be generalized to larger groups of examiners and subjects.

The present experiment attempted to overcome this weakness in the independent variable by using eight male graduate students in psychology as examiners. Each examiner had completed a course in individual intelligence test administration and scoring, in addition to having had experience along these lines in a practicum setting.

Materials

Intelligence Test:

Two series of five subtests (Pentads) each from the Wechsler Adult Intelligence Scale were used as measures of intelligence test performance. Pentad I consisted of subtests: Information, Similarities, Vocabulary, Picture Arrangement, and Object Assembly. Pentad II consisted of subtests: Comprehension, Arithmetic, Digit Symbol, Picture Completion and Block Design.

According to Maxwell (1957), who followed up the work of McNemar (1950) and Doppelt (1956) on estimating full scale IQ from short forms of the WAIS, Pentad I correlates .972 with full scale score and Pentad II correlates .966 with full scale score. These correlations

receive indirect support from the research of Howard (1959) and Clayton and Payne (1959) who tested the validity of similar short forms.

Thus these two Pentads estimate with little error variance the measurement of intelligence as shown by the full scale WAIS. Because each fails to account for approximately .06 of the total variance, these two Pentads could, at a very minimum, correlate .94 with each other. On this basis it was felt that the two Pentads could be regarded as equivalent enough to compare the effects of the treatment conditions. Standard WAIS record forms were used to record the data.

Rating Scale:

The rating scale consisted of twenty-seven personal adjectives representative of four major areas of personality (Dominance, Activity, Social Sensitivity, and Mood) which, it was felt, is pertinent to psychological testing. The rating was along a continuum of Strongly Agree to Strongly Disagree.

The rating scale was derived from the SAQS Chicago Q Sort (Corsini, 1956), which seemed to provide a valuable research tool for this experiment, because it contains a number of previously selected personal adjectives describing people. These adjectives are applicable for the study of the interaction between the examiner and the subject. Table 1 presents the actual rating scale and Table 5, the adjectives grouped according to major personality areas and a comparison of the subject's ratings of the examiners under the two different treatment conditions.

TABLE 1
RATING SCALE

Name _____

The following words are descriptions of people. You are to describe your _____ Examiner by encircling the letters which signify whether you Strongly Agree (SA), Agree (A), Undecided (U), Disagree (D), and Strongly Disagree (SD), with the descriptive word. The results of the Rating Scale will be held in strict confidence, so please be frank.

1. Aggressive	SA	A	U	D	SD
2. Quick	SA	A	U	D	SD
3. Warm-hearted	SA	A	U	D	SD
4. Easy-going	SA	A	U	D	SD
5. Forceful	SA	A	U	D	SD
6. Hasty	SA	A	U	D	SD
7. Soft-hearted	SA	A	U	D	SD
8. Calm	SA	A	U	D	SD
9. Independent	SA	A	U	D	SD
10. Hurried	SA	A	U	D	SD
11. Gentle	SA	A	U	D	SD
12. Worrying	SA	A	U	D	SD
13. Stubborn	SA	A	U	D	SD
14. Talkative	SA	A	U	D	SD
15. Appreciative	SA	A	U	D	SD
16. Emotional	SA	A	U	D	SD
17. Dominant	SA	A	U	D	SD
18. Active	SA	A	U	D	SD

TABLE 1--(Continued)

19.	Discreet	SA	A	U	D	SD
20.	Excitable	SA	A	U	D	SD
21.	Outspoken	SA	A	U	D	SD
22.	Unselfish	SA	A	U	D	SD
23.	Submissive	SA	A	U	D	SD
24.	Insensitive	SA	A	U	D	SD
25.	Dependent	SA	A	U	D	SD
26.	Sensitive	SA	A	U	D	SD
27.	Sarcastic	SA	A	U	D	SD

Negative and Positive Treatment Conditions

In order to evaluate adequately the results of this experiment, it was important to develop some degree of standardization in the administration of the negative and positive treatment conditions. Since the greatest part of the administration consists of a verbal interaction between the examiner and subject, procedures were developed to create a difference between positive and negative treatments in terms of the manner of administration rather than in terms of a change in the actual directions of administration which appear in the WAIS manual (Wechsler, 1955). It was also felt that this approach would generalize the results to the actual day-by-day professional use of the WAIS.

The treatment conditions are defined in the following instructions to examiners.

Positive Administration: In general, this administration is characterized by a warm and highly approving, interested manner of test administration. The examiner is prone to give reinforcing statements in a warm tone of voice, i.e., he will be personally warm and appreciative in manner. The examiner will speak in an encouraging tone of voice and look at the subject with a smile while asking questions, preparing tests, or giving directions. In other words, he will make every "Hm!" sound like a compliment for work well done. The specific activities of the examiner are:

1. Introduce yourself to the subject and shake hands with him.
2. Look at the subject while talking to him.
3. Use a tone of voice which reflects warmth and interest, e.g., speak slowly, softly, and clearly and vary the intonation pattern.
4. Demonstrate appreciation of the subject's responses by stating such things as "good," "you're doing O.K.," etc.
5. Wait patiently for each response.
6. Appear alert to everything that the subject says, i.e., avoid appearing bored or tired.

7. Be quick to smile at appropriate instances, e.g., for a good response or when the subject smiles.
8. When a subtest is finished, the transition to the next subtest should be facilitated by such remarks as, "Here is something of a different sort" or "I think that you will find this interesting." Never go from one subtest to another without saying something of a positive and rewarding nature.

Negative Administration: In general, this administration is characterized by a persistently rejecting and disinterested manner of test administration. The examiner is prone to make punishing statements, consisting of remarks and actions designed to insult the subject or the response made, i.e., the examiner will assume the role of a harsh, rejecting, authoritarian figure. He will be deliberately unconcerned about the subject and will not look at him while asking questions, preparing tests, or giving directions; he will never smile and will give directions in a voice of dictatorial harshness, making every "Hm!" sound like a sneer. The specific activities of the examiner are:

1. Do not introduce yourself to the subject and do not acknowledge his attempts at introduction.
2. Never look at the subject while talking to him.
3. Use a tone of voice which reflects coldness and disinterest, e.g., speak rapidly but clearly in a steady monotone with no variance of the intonation pattern.
4. Demonstrate rejection and lack of appreciation by frowning and not saying anything while the response is being given.
5. Demonstrate impatience by resorting to such things as tapping the table with your pencil, looking up at the ceiling, and becoming fidgety as the subject responds.
6. Create the impression of being bored by sighing heavily upon occasion and manhandling the test materials in order to create an impression of distaste for what is being done.
7. Do not smile and do not respond to the subject's attempts to be pleasant during the testing.
8. When a subtest is finished, the transition to the next subtest should be characterized by immediately launching into the administration with no preliminary remarks other than a disdainful grunt.

Specific application of these instructions to the various subtests are:

Pentad I: Information, Similarities, and Vocabulary are mainly tests involving a verbal interaction between the subject and the examiner. The above suggestions for Positive and Negative treatments should be applied throughout the testing.

With Picture Arrangement and Object Assembly, there is an opportunity, through the actual manipulation of test materials, to delineate further the difference between Positive and Negative treatments. In the Positive treatment the examiner presents the materials as they are normally presented. In the Negative treatment the examiner presents the materials by forcefully putting them on the table and removing them in the same manner.

Pentad II: Comprehension and Arithmetic are to be handled the same as Information, Similarities, and Vocabulary are handled in Pentad I for the Positive and Negative treatment.

Digit Symbol, Picture Completion, and Block Design are to be handled in the same manner as Picture Arrangement and Object Assembly are handled in Pentad I for the Positive and Negative treatment.

Procedure

To reduce the effects of order and sequence which introduce confounding errors of measurement, the experiment followed a pattern of counterbalancing.

The eight examiners were broken down into four blocks with two examiners in a block. Within each block there were twelve subjects selected proportionately from the A.C.E. divisions discussed earlier. All of the subjects acted as their own controls; thus they experienced both treatment conditions, but each from a different examiner within a block.

Since possible sources of error could develop from the examiners administering one form of the treatment condition, the examiners

alternated between negative and positive administrations, so that each examiner gave a total of six negative administrations and six positive administrations with a pattern of alternating from one treatment condition to another throughout the sequence of twelve administrations. To control further any possible sequence effect, the subjects were assigned on a counterbalanced basis to various examiners and treatment conditions. More specifically, a subject who received an administration early in that examiner's sequence did not receive his second administration until late in another examiner's sequence.

To control for the Pentads themselves introducing error by being used consistently in one treatment condition, the Pentads were counterbalanced with regard to the use of one Pentad with one treatment condition. Thus the Pentads were used as negative and positive treatment stimuli equally.

The experiment was run in the evening, each examiner working in a separate office testing three subjects each evening for a total of four evenings. The subjects each appeared twice on consecutive evenings.

All of the subjects were naïve about the purpose of the experiment. They were told that the experiment was an investigation in the area of psychological testing and that their part in the experiment would be to answer a list of standardized questions.

At the end of the second testing session the subjects were given two copies of the rating scale appearing in Table 1. The subjects were asked to rate their "first examiner" and their "second examiner" in order to maintain subject naïveté during this last phase of the experiment. The subjects were then asked not to discuss the experiment with

anyone until they were informed that the experiment was over.

Scale scores for the subtests were used as data in comparing the effects of the treatments. The Pentads were scored by the experimenter using the recommended scoring which appears in the WAIS manual.

CHAPTER III

RESULTS AND DISCUSSION

Results

The design of this experiment allowed an analysis of variance to be done on the subtest scale score totals for the Pentads under the different treatment conditions. Table 2 presents the results of the analysis of variance. The missing subject discussed earlier was handled by replacing his missing score by a value equal to the mean of the other scores in that subject's treatment-cell and subtracting one degree of freedom from the degrees of freedom for total and consequently from the degrees of freedom for error also.

The insignificant F for the variance between examiners indicates a homogeneity between examiners suggestive of their handling the treatment conditions in a very similar manner. Table 3 suggests this, because a high degree of correspondence exists for subtest totals from block to block for both treatments.

The insignificant interaction between examiners and treatments can also be interpreted along the lines of examiner similarity in handling the different treatment conditions. The main implication here is that the treatments had the same relative effects throughout the blocks or levels.

The F of 2.86 for the variance between negative and positive treatments falls short of the F of about 4.0 required for significance

at the .05 level. This F approaches significance, and there are probably many reasons for a significant difference not appearing at the present time. This F is probably the most important phase of the statistical findings, and it tends to support a hypothesis that the attitude of the examiner has little effect on test results in well-structured testing. Variables to be discussed later not included or manipulated in this experiment suggest that this hypothesis may be unwarranted.

To present the direction of the shift of scores from one treatment condition to the other and to determine general consistency of this shift, the estimated average Full Scale IQ per block of examiners for both treatment conditions is listed in Table 3. It is apparent that the shift represents a consistent decrement in score in the negative treatment condition. Again there is the suggestion that the Pentads are equivalent and have little to do with confounding the results because of the counterbalancing of the Pentads. Although examiners 3 and 4 experience a mild shift while examiners 7 and 8 experience a more severe shift, it is observed that the difference in shift is not intense enough to create a significant F for interaction.

It was felt that an examination of the effect of the treatments on the individual subtests would be of interest and value. Table 4 lists the results of a series of t tests done on each subtest comparing scale scores for both treatment groups.

A general decrement in score occurs within the negative treatment group for Information, Comprehension, Similarities, Vocabulary, Block Design, and Picture Arrangement. A slight increase in score for the negative treatment appears for Digit Symbol, Picture Completion,

and Object Assembly, while Arithmetic demonstrates no change at all. These differences are insignificant except for Vocabulary, which is significant at the .02 level of significance. Why this particular sub-test should demonstrate significance will be discussed later.

Table 5 presents a comparison of the direction of the subject's ratings of the examiners under the positive and negative treatment conditions using the rating scale presented in Table 1. Table 5 also groups the descriptive adjectives used for the ratings according to major personality areas.

The comparisons of whether or not the subjects agreed with the adjective as being descriptive of the examiner were made using the chi-square technique as a measure of the significance of difference of the agree-disagree ratings under each treatment condition. Those adjectives which were found to be significant are checked in order to indicate the direction of the rating under each treatment condition.

Of the twenty-seven adjectives it is noticed that eighteen achieve a significant level of difference in ratings of the examiners. It is probably noteworthy that under the personality area of Social Sensitivity all nine of the adjectives are significant, whereas the area of Dominance yields only three out of eight significant differences, Activity yields three out of five, and Mood also yields three out of five significant differences.

More specifically, examiners under the positive treatment are seen by the subjects as talkative, warm-hearted, gentle, appreciative, discreet, unselfish, sensitive, easy-going, and calm. The same examiners under the negative treatment condition are not seen as possessing

these qualities. The negative treatment produced a picture of the examiners as being forceful, stubborn, dominant, hasty, hurried, insensitive, sarcastic, and worrying. Conversely, the same examiners under the positive treatment do not receive these ratings.

TABLE 2
ANALYSIS OF VARIANCE

Source	DF	MS	F	P
E: Examiners	3	5.97	< 1	Not sig.
T: Treatments	1	130.66	2.86	> .05
E x T	3	30.19	< 1	Not sig.
Error Term	<u>87</u>	45.73		
Total	94			

TABLE 3
ESTIMATED AVERAGE FULL SCALE IQ's FOR
POSITIVE AND NEGATIVE TREATMENTS

Blocks	Positive Treatment		Negative Treatment	
	Pentad	Average IQ	Pentad	Average IQ
1 and 2	1	115	2	114
3 and 4	2	114	1	114
5 and 6	1	117	2	113
7 and 8	2	118	1	111

TABLE 4
INDIVIDUAL SUBTEST SCALE SCORES UNDER
EACH TREATMENT CONDITION

Subtest	Condition	Mean	Mean Difference	SD	t	P
Information	Positive	13.35	.35	1.60	1.02	Not sig.
	Negative	13.00				
Comprehension	Positive	13.35	1.00	4.15	1.13	Not sig.
	Negative	12.35				
Arithmetic	Positive	12.42	0	3.31	0	Not sig.
	Negative	12.42				
Similarities	Positive	13.25	1.00	3.09	1.51	Not sig.
	Negative	12.25				
Vocabulary	Positive	13.43	1.35	2.39	2.65	.02
	Negative	12.08				
Digit Symbol	Positive	11.69	- .13	2.95	- .21	Not sig.
	Negative	11.82				
Picture Completion	Positive	12.07	-.39	2.80	- .66	Not sig.
	Negative	12.46				
Block Design	Positive	12.13	.56	4.26	.58	Not sig.
	Negative	11.57				
Picture Arrangement	Positive	10.96	.30	2.44	.51	Not sig.
	Negative	10.66				
Object Assembly	Positive	10.17	- .87	4.32	- .94	Not sig.
	Negative	11.04				

TABLE 5

COMPARISON OF SUBJECTS' RATINGS OF EXAMINERS
UNDER POSITIVE AND NEGATIVE TREATMENTS

Adjective	DF	x ²	P	Positive Treatment		Negative Treatment	
				Agree	Disagree	Agree	Disagree
<u>DOMINANCE</u>							
Aggressive	3	.88	<.80				
Forceful	4	12.36	<.01		x	x	
Independent	3	3.92	<.50				
Stubborn	3	25.90	<.01		x	x	
Dominant	3	17.56	<.01		x	x	
Outspoken	4	3.54	<.50				
Submissive	4	6.98	>.10				
Dependent	3	.82	<.90				
<u>ACTIVITY</u>							
Quick	3	.24	<.98				
Hasty	3	28.92	>.01		x	x	
Hurried	3	31.20	>.01		x	x	
Talkative	3	46.02	>.01	x			x
Active	3	2.80	<.50				
<u>SOCIAL</u>							
<u>SENSITIVITY</u>							
Warm-hearted	3	48.27	>.01	x			x
Soft-hearted	3	29.04	>.01	x			x
Gentle	3	44.25	>.01	x			x
Appreciative	3	38.12	>.01	x			x
Discreet	4	10.52	>.05	x			x
Unselfish	3	21.90	>.01	x			x
Insensitive	3	21.88	>.01		x	x	
Sensitive	3	11.28	.01	x			x
Sarcastic	3	24.06	>.01		x	x	
<u>MOOD</u>							
Easy-going	4	44.92	>.01	x			x
Calm	3	24.80	>.01	x			x
Worrying	3	29.12	>.01		x	x	
Emotional	4	4.06	<.30				
Excitable	4	8.72	>.10				

Discussion

With regard to Hypothesis I--that there will be a consistent shift in intelligence test performance as a result of the different experimentally established negative and positive interactions between the examiner and the subject--little support is received by the results of this study. There was a consistent shift, but it was at an insignificant level of significance ($P > .05$).

Perhaps different results could be attained by utilizing a different sampling procedure. Studying class attitudes towards psychiatry, Redlich, Hollingshead, and Bellis (1955) found that there was a higher degree of rapproachment between the value systems of psychotherapists and of middle-class patients than between the value systems of psychotherapists and lower-class patients. Evidence for this conclusion was found in the facts that lower-class patients did not enter therapy voluntarily, the therapists demonstrated a greater dislike of lower-class patients than of middle-class patients, and poor communication existed between therapists and lower-class patients. The sample used in this study, for both examiners and subjects, were very much alike in class attitude and values because of their identification with the academic atmosphere.

By generalizing from the Redlich et al. study to motivational factors and the establishment of rapport in intelligence testing, it is possible to suggest different sets of variables coming into play with such groups as maladjusted individuals, minority groups, or persons with different social norms and values. If these variables were taken into consideration and manipulated in an experiment similar to the present study they could well demonstrate their importance by being reflected in more significant results.

Another possibility is for a related study in which the personalities of the examiners and subjects are paramount. Although the experimental difficulties in such a study are, of course, great--especially in respect to evaluating the examiner's personality--this interesting problem merits consideration.

Hypothesis II--that the shift in performance will most likely reflect a decrement in score for the majority of the subjects under the negative treatment condition--receives support, although not significant support, from the results. No explanation of these results can be derived from this experiment. A possible explanation may be found in related research dealing with the detrimental effects of anxiety and tension upon intelligence test performance. Sarason et al. (1952) found that, when subjects with a low-anxiety rating were given ego-involving instructions on a stylus maze, a slight increase in performance resulted. Subjects with a high-anxiety rating, however, did poorly under ego-involving instructions. Wiener (1957), using "distrustfulness" and "suspiciousness" as independent variables, found that subjects high in these traits tend to get lower IQ's because they hold back full answers or actually deny the implications of test questions.

The suggestion from such related research is that the "anxiety" and "distrust" aroused by the negative treatment condition tends to lower intelligence test scores. Again there are implications for further research within the present experimental framework.

The results of the rating scale support both Hypothesis III--that the subject's ratings of the examiners will be clearly different as a result of the different treatment conditions--and Hypothesis IV--

that subjects will perceive the negative-treatment examiners in an unfavorable light, whereas they will perceive the positive-treatment examiners in a favorable light. These results also indirectly support the experimental reality of the two treatment conditions in that the same examiners received different ratings in accordance with the treatment conditions.

As a means of studying the effect and operating traits of the two treatment conditions, the rating scale does an adequate job. It is a step in the direction of studying the actual interaction between the examiner and the subject. This study was concerned mainly with experimentally established interpersonal variables influencing test performance and not with an attempt to determine how these variables actually affect the subject. Such a study would require an analysis of the interaction process itself. In the area of research in psychotherapy such techniques are being developed (Rubinstein and Parloff, 1959).

Of the individual subtests, Vocabulary was the only one demonstrating a significant decrement in score as a result of the negative condition. The mental processes involved in this subtest are generally regarded as the ability to recall previously acquired verbal meanings with variations in the quality of responses yielded. Perhaps these processes are upset by negative influences; in other words, the quality of recall and response may suffer as a result of the negative treatment condition.

Although the results of this experiment do not significantly demonstrate a decrement in intelligence test scores, the other results

of the experiment and related areas of experimentation discussed in the previous paragraphs suggest further research along this line with structured psychological tests. The assumption that structured tests are not affected by interpersonal and situational variables is still questionable.

If examiner-subject relationships can be demonstrated with structured tests, the interpretation of test results must go beyond a purely psychometric interpretation. Schafer suggests that the testing situation be looked upon as a social situation from which to derive data in order to understand the subject better. This approach to testing necessarily increases the possibility of personalized interpretation, but it also allows for a greater inclusion of data. For this "total-situation" approach to be admissible to psychological testing, a system of handling accurately great complexities of data must be worked out, in order to satisfy the criterion of objectivity in test interpretation.

CHAPTER IV

SUMMARY

The purpose of this study was to determine whether or not experimentally established examiner attitudes--negative and positive--could influence intelligence test scores. The study was also concerned with identifying examiner traits of the two treatment conditions.

The subjects used in this study were forty-eight male undergraduate students, who acted as their own controls by participating in both the negative and positive treatment conditions. The examiners were eight male graduate students in psychology sophisticated in intelligence test administration. The examiners received specific instructions concerning their behavior during the administration of the treatment conditions.

The measures of intelligence were two short forms of the Wechsler Adult Intelligence Scale. These short forms were composed of five subtests each and correlated approximately .97 with full scale WAIS and at least .94 with each other. Selected parts of the Chicago Q Sort were used as a rating scale of the examiners.

The experiment followed a pattern of counterbalancing in order to control for order and sequence effects. Each examiner tested twelve subjects and alternated negative and positive administrations. Each subject received a negative and positive administration from a different examiner and alternated positions within the examiner's sequence.

The Pentads received equal use in both treatment conditions in order to control for biasing factors arising from Pentad differences. At the end of the second testing session the subjects filled out a rating scale of the two examiners who worked with him in the experiment.

Scale scores for the subtests were used as data in comparing the effects of the treatments. An analysis of variance of this data revealed no significant F's other than the variance between treatments approaching significance. It was felt that the utilization of different types of subject populations, other than college sophomores, might well yield more significant results. The majority of the scores decreased as a result of the negative treatment condition.

An analysis of the subtests separately revealed that Vocabulary was the one subtest which demonstrated a significant decrease as a result of the negative treatment condition.

Chi-squares of the rating scale data demonstrated clear differences in the subjects' ratings of the same examiners in the different treatment conditions.

It was felt that more research with well-structured tests, e.g., intelligence tests, with regard to interpersonal and situational influences should be done. It was urged that psychologists look upon the testing situation as a social situation capable of delivering more data than a pure psychometric approach could yield. The problem of personalized interpretation was pointed out and it was suggested that a system for accurately and objectively handling such complexities be evolved.

BIBLIOGRAPHY

- Bernstein, L. The examiner as inhibiting factor in clinical testing. J. consult. Psychol., 1956, 20, 287-290.
- Clayton, H. and Payne, D. Validation of Doppelt's WAIS short form with a clinical population. J. consult. Psychol., 1959, 23, 467.
- Corsini, R. J. SAQS Chicago Q Sort. Chicago: Psychometric Affiliates, 1956.
- Cronbach, L. J. Essentials of psychological testing (2nd ed.). New York: Harper and Brothers, 1960.
- Doppelt, J. Estimating the Full Scale Score on the Wechsler Adult Intelligence Scale from scores on four subtests. J. consult. Psychol., 1956, 20, 63-66.
- Gibbey, R. G., Miller, D. R., and Walker, E. L. The examiner's influence on the Rorschach protocol. J. consult. Psychol., 1953, 17, 425-428.
- Gross, L. Effects of verbal and non-verbal reinforcement in the Rorschach. J. consult. Psychol., 1959, 23, 66-68.
- Hammond, K. R. Representative vs. systematic design in clinical psychology. Psychol. Bull., 1954, 51, 150-159.
- Henry, Edith and Rotter, J. B. Situational influence on Rorschach responses. J. consult. Psychol., 1956, 6, 457-462.
- Howard, W. Validities of WAIS short forms in a psychiatric population. J. consult. Psychol., 1959, 23, 282.
- Hutt, M. L. "Consecutive" and "adaptive" testing with the revised Stanford-Binet. J. consult. Psychol., 1947, 11, 93-104.
- Joel, W. The interpersonal equation in projective methods. Rorschach Res. Exch., 1949, 13, 479-482.
- Lantz, B. Some dynamic aspects of success and failure. Psychol. Monogr., 1945, 59, 6-21.
- Lord, Edith. Experimentally induced variations in Rorschach performance. Psychol. Monogr., 1950, 64 (10, Whole No. 316).
- Luft, J. Interaction and projection. J. proj. Tech., 1953, 17, 489-492.

- McNemar, Q. An abbreviated Wechsler-Bellevue scale. J. consult. Psychol., 1950, 14, 79-81.
- Masling, J. M. The effects of warm and cold interaction on the interpretation of a projective protocol. J. proj. Tech., 1957, 21, 377-383.
- _____. The effects of warm and cold interaction on the administration and scoring of an intelligence test. J. consult. Psychol., 1959, 23, 336-339.
- _____. The influence of situational and interpersonal variables in projective testing. Psychol. Bull., 1960, 57, 65-85.
- Maxwell, Eileen. Validities of abbreviated WAIS scales. J. consult. Psychol., 1957, 21, 121-126.
- Rabin, A., Nelson, W., and Clark, Margaret. Rorschach content as a function of perceptual experience and sex of the examiner. J. clin. Psychol., 1954, 10, 188-190.
- Redlich, F. C., Hollingshead, A. B., and Bellis, Elizabeth. Social class differences in attitudes toward psychiatry. Amer. J. Orthopsychiat., 1955, 25, 60-70.
- Rubinstein, E. A. and Parloff, M. B. (eds.) Research in psychotherapy. Washington: American Psychological Association, 1959.
- Sacks, E. L. Intelligence scores as a function of experimentally established social relationships between child and examiner. J. abnorm. soc. Psychol., 1952, 47, 354-358.
- Sanders, R. and Cleveland, S. E. The relationship between certain examiner personality variables and subjects' Rorschach scores. J. proj. Tech., 1953, 17, 34-50.
- Sarason, S. B., Mandler, G., and Craighill, P. C. The effect of differential instructions on anxiety and learning. J. abnorm. soc. Psychol., 1952, 47, 561-565.
- Schafer, Roy. Psychoanalytic interpretation in Rorschach testing. New York: Grune and Stratton, 1954.
- Staudt, Virginia M. The relationship of testing conditions and intellectual level to errors and correct responses in several types of tasks among college women. J. Psychol., 1948, 26, 125-140.
- Wechsler, D. Manual for the Wechsler Adult Intelligence Scale. New York: Psychological Corp., 1955.

Wickes, T. A. Examiner influence in a test situation. J. consult. Psychol., 1956, 20, 23-26.

Wiener, G. The effect of distrust on some aspects of intelligence test behavior. J. consult. Psychol., 1957, 21, 127-130.

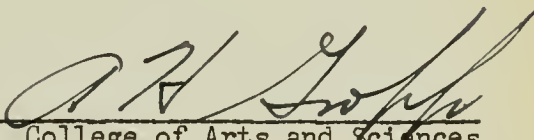
BIOGRAPHICAL SKETCH

William George Murdy, Jr., was born in New York City, New York, on June 5, 1934. In June, 1952, he was graduated from Andrew Jackson High School, St. Albans, Long Island. As an undergraduate, he studied his first two years at the University of Florida and then transferred to New York University, where he received, in June, 1956, the degree of Bachelor of Arts with majors in English and Psychology. In September, 1956, he was admitted to the University of Florida as a graduate student in Psychology. The following year he entered the United States Army for six months.

After his marriage to Thelma Louise Baughan in August, 1958, he resumed his graduate studies at the University of Florida. In June, 1959, he received the degree of Master of Arts and began working toward the doctoral degree. As a trainee, he joined the Veterans Administration Clinical Psychology Training Program in Gulfport, Mississippi, in March, 1961. Since that time, he has been completing the academic requirements and internship necessary for specialization in clinical psychology.

This dissertation was prepared under the direction of the chairman of the candidate's supervisory committee and has been approved by all members of that committee. It was submitted to the Dean of the College of Arts and Sciences and to the Graduate Council, and was approved as partial fulfillment of the requirements for the degree of Doctor of Philosophy.

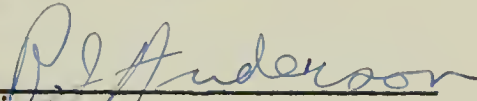
February 3, 1962



Dean, College of Arts and Sciences

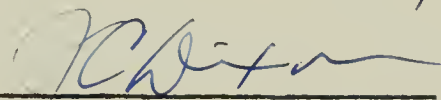
Dean, Graduate School

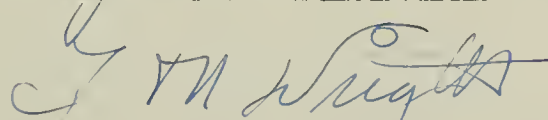
SUPERVISORY COMMITTEE:

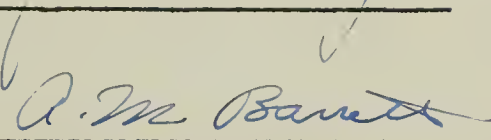


Chairman









¹
32847 3^v